PACIFIC JOURNAL OF MATHEMATICS
Vol. 109, No. 2, 1983

BERNSTEIN-LIKE POLYNOMIAL APPROXIMATION IN HIGHER DIMENSIONS

LESTER E. DUBINS

Let $C(K)$ be the Banach space of continuous, real-valued functions defined on a compact, Hausorff space, K, let $\mathcal{P} = \mathcal{P}(K)$ be the positive linear forms, P, defined on $C(K)$, for which $Pf \leq \sup f(k)$ $(k \in K)$, $f \in C(K)$, and endow $\mathcal{P}(K)$ with the weak-star topology in which it is, of course, compact. (As is well known, $\mathcal{P}$ can be identified with the set of countably additive probability measures defined on the Baire subsets of K.) Let P^∞ be the power probability on K^∞, the product of denumerable number of copies of K. Then, for each $f \in C(K^\infty)$, the integral of f with respect to P^∞ is plainly continuous in P. As Theorem 2 below states, there are no ohter continuous real-valued functions of P. The proof of this assertion requires a generalization of Bernstein's version of the celebrated polynomial approximation theorem of Weierstrass, which generalization is provided by Theorem 1.

The two theorems are numbered in their logical order but stated in the order of simplicity of formulation.

THEOREM 2. *For every $g \in C(\mathcal{P}(K))$, there is an $f \in C(K^\infty)$ such that, for all $P \in \mathcal{P}(K)$,*

$$\int f\,dP^\infty = g(P). \tag{1}$$

It is convenient to reformulate Theorem 2 in terms of an operator T mapping $C(K^\infty)$ into $C(\mathcal{P}(K))$ defined, thus.

$$(Tf)P = \int f\,dP^\infty, \qquad f \in C(K^\infty), P \in \mathcal{P}(K). \tag{2}$$

Reformulated, Theorem 2, states that the operator T is surjective, that is, onto $C(\mathcal{P})$. A short digression explains the origin of this theorem.

Suppose Q is a probabilistic mixture of power probabilities,

$$Q = \int P^\infty \mu(dP) \tag{3}$$

for some probability μ on $\mathcal{P}$, or, more fully,

$$Qf = \int (Tf)(P)\,\mu(dP), \qquad f \in C(K). \tag{4}$$

As a theorem of Hewitt and Savage [2] asserts, each Q is representable in at most one way as a mixture of power probabilities. Plainly, for the uniqueness of μ, it is sufficient that every $g \in C(\mathcal{P})$ be in the uniform closure of functions of the form Tf for $f \in C(K)$. It was this observation which led me to investigate whether even the stronger assertion of Theorem 2 might be true, which explains the genesis of this paper.

To Benjamin Weiss[1] I am quite indebted. For it was he who, in a private communication to me, proved Theorem 2 when K is a two-point set. Moreover, his argument carries over to arbitrary compact sets, K, once Bernstein's polynomial approximation theorem is generalized, as in Theorem 1 below.

Recall that Bernstein has shown that, for each continuous function, g, defined on the closed unit interval, the polynomials g_n converge uniformly to g where

$$g_n(p) = \sum_{j=1}^{n} \binom{n}{j} g\left(\frac{j}{n}\right) p^j (1-p)^{n-j}. \tag{5}$$

To state the generalization of Bernstein's theorem, introduce the usual notation, δ_k, for the probability supported by the singleton $\{k\}$, that is, $\delta_k \in \mathcal{P}(K)$ is the evaluation map: $\phi \to \phi(k)$, for $\phi \in C(K)$. Letting $\omega = (k(1), k(2), \ldots)$ be a generic point of K^∞, introduce $D_n(\omega)$, its empirical distribution of order n,

$$D_n(k(1), k(2), \ldots) = \frac{1}{n} \sum_{i=1}^{n} \delta_{k(i)}. \tag{6}$$

Of course, $D_n\colon K^\infty \to \mathcal{P}$ is continuous. Therefore, its composition with a $g \in C(\mathcal{P})$, $g \circ D_n$, also to be designated by $S_n g$, is a continuous, real-valued mapping with domain K^∞ which has an expectation under any probability on K^∞, in particular, under any power probability P^∞. Recapitulating, $S_n\colon C(\mathcal{P}) \to C(K^\infty)$ and $TS_n\colon C(\mathcal{P}) \to C(\mathcal{P})$ where T is defined as in (2). In fuller detail,

$$(TS_n g)P = \int S_n g \, dP^\infty = \int g \circ D_n \, dP^\infty, \tag{7}$$

for $g \in C(\mathcal{P})$, $P \in \mathcal{P}$.

THEOREM 1. *TS_n converges in the strong operator topology to the identity on $C(\mathcal{P})$, that is, for each $g \in C(\mathcal{P})$, (7) converges to $g(P)$ uniformly in P.*

[1]Rather than being a coauthor, Weiss invited me to use the ideas and the argument contained in his communication.

Of course, if K is a two-point set $\{a, b\}$, $\mathcal{P}$ is identified with the closed unit interval via the correspondence $P \Leftrightarrow p$ if, and only if, $P(b) = p$, and, in this case, $TS_n g$ is the Bernstein polynomial g_n of (5). If K is a finite set, $\mathcal{P}$ is a simplex and one obtains from Theorem 1 similar polynomial approximation to g. For instance, if K is a three-point set, Δ is the triangle of all triplets (p, q, r) of nonnegative numbers whose sum is 1, $g \in C(\Delta)$.

$$g_n(p, q, r) = \sum \binom{n}{i, j, k} g\left(\frac{i}{n}, \frac{j}{n}, \frac{k}{n}\right) p^i q^j r^k, \tag{8}$$

where the sum is over all triplets of nonnegative integers whose sum is n, and where the familiar multinomial notation is being used; and g_n converges to g uniformly on Δ.

It is an immediate corollary of Theorem 1 that the range of T is dense in $C(\mathcal{P})$ which, of course, yields another proof of the Hewitt and Savage [**2**] result which asserts the uniqueness of the representation of an exchangeable probability as a mixture of power probabilities.

2. Quadratic functions. If Y is a topological linear space, later to be specialized to be the dual of $C(K)$ in the usual weak-star topology, a finite linear combination of functions, each of which is either a constant, a continuous linear functional defined on Y, or the square of such a linear function, is a *finite-rank quadratic function*. The restriction of a quadratic function to a subset of Y is called a quadratic function on that subset. Throughout this note, Y is endowed with the weakest topology which permits each element x of a linear space X of linear functionals on Y to be continuous.

LEMMA 1. *Let $y \in V \subset Y$, where V is a neighborhood of y, and let b and c be real numbers. Then there is a finite-rank quadratic function q: $Y \to \mathbf{R}$ such that $q(y) = b$, $q \geq b$ everywhere, and $q \geq c$ on the complement of V.*

Proof of Lemma 1. Since the space of finite-rank quadratic functions is invariant under translations of Y, as well as under positive affine transformations of $\mathbf{R}$, it suffices to treat the case where y is the origin of Y, $b = 0$, and $c = 1$. Since Y has the weak topology induced by X, there exist $x_i \in X$, $1 \leq i \leq n$, such that $|(x_i, z)| < 1$ for all i implies $z \in V$. Plainly, for $q(z) = \sum (x_i, z)^2$, $q(0) = 0$, $q \geq 0$ everywhere, and $q \geq 1$ on the complement of V. □

LEMMA 2. *Let g be a bounded function defined on a subset L of Y which, at a point $y \in L$, is continuous (or even semicontinuous). Then, for every $\varepsilon > 0$, there is a finite-rank quadratic function q on L which majorizes g and which, at y, is $g(y) + \varepsilon$.*

Proof. Let V be a neighborhood of y such that, on V, g is nowhere greater than $g(y) + \varepsilon$. The preceding lemma applies with $b = g(y) + \varepsilon$ and c equal to any upper bound for g. □

Let J be the set of finite-rank quadratic functions on L.

LEMMA 3. *Let $g \in C(L)$ where L is a compact subset of Y. Then*

$$g(y) = \inf\{q(y) | q \geq g, q \in J\}, \tag{1}$$

and

$$g(y) = \sup\{q(y) | q \leq g, q \in J\}. \tag{2}$$

Proof. From Lemma 2, (1) is immediate, and (2) follows by applying (1) to $-g$. □

PROPOSITION 1. *Let L be a compact subset of Y, and let $T_1, T_2, \ldots$ be a sequence of order-preserving (not necessarily linear) mappings of $C(L)$ into itself such that $T_n q$ converges to q whenever q is a finite-rank quadratic function. Then T_n converges in the strong operator topology to the identity operator, that is, for each $g \in C(L)$, $T_n g$ converges to g uniformly on L.*

Proof. Let $g \in C(L)$. Then as Lemma 3 asserts, g satisfies (1) and (2). And, as shown by Bauer [**1**, Proposition 1] which in turn was inspired by results of Korovkin [**3**], for every g which satisfies (1) and (2), $T_n g \to g$ uniformly on L. This completes the proof, but for the convenience of the reader, here is a sketch of the argument that $T_n g \to g$. First, use (1) and (2), and the compactness of L to verify that, for every $\varepsilon > 0$, there exist $q_i \in J$, $1 \leq i \leq m$, each of which majorizes g, but whose infimum is majorized by $g + \varepsilon$. By hypothesis, $\exists N$ such that $T_n q_i < q_i + \varepsilon$ for $n \geq N$ and all i. Since T_n is order-preserving, one obtains

$$T_n g < \inf_i T_n q_i < \inf_i q_i + \varepsilon < g + 2\varepsilon. \tag{3}$$

A similar calculation shows that, for $n \geq N'$, $T_n g$ exceeds $g - 2\varepsilon$. □

3. Proof of Theorem 1. Fix $\phi \in C(K)$ and let

$$v_\phi P = \int \phi^2\, dP - \left(\int \phi\, dP\right)^2, \tag{1}$$

which is often called the variance of ϕ with respect to P, and let

$$q_\phi P = \left(\int \phi\, dP\right)^2. \tag{2}$$

LEMMA 1. $TS_n q_\phi = q_\phi + \frac{1}{n} V_\phi$.

The following straightforward calculation comprises the proof of Lemma 1.

$$\begin{aligned}(TS_n q_\phi)P &= \int q_\phi \circ D_n\, dP^\infty \\ &= \int \left(\frac{1}{n}\sum \phi(k_i)\right)^2 dP^\infty \\ &= \frac{1}{n^2}\left\{\int \sum_{i\neq j} \phi(k_i)\phi(k_j) + \sum \phi^2(k_i)\right\} dP^\infty \\ &= \int \phi(k_1)\phi(k_2)\, dP^2 + \frac{1}{n}\left\{\int \phi^2\, dP - \int \phi(k_1)\phi(k_2)\, dP^2\right\} \\ &= \left(\int \phi\, dP\right)^2 + \frac{1}{n}\left(\int \phi^2\, dP - \left(\int \phi\, dP\right)^2\right) \qquad \square\end{aligned} \tag{3}$$

(Incidentally, with the exception of the last equality, all equalities obtain if P^∞ were any exchangeable probability or, more generally, any second-order exchangeable probability.)

Let Y be the dual of $C(K)$ and endow Y with the weak-star topology. As is well known, functions of the form $P \to \int \phi\, dP$ for some $\phi \in C(K)$ are the only continuous linear functionals on Y. As is verified without difficulty, if g is such a function, or is constant, then $TS_n g = g$ for all n. This, together with Lemma 1, implies that $TS_n g \to g$ uniformly on $\mathcal{P}$ whenever g is a finite-rank quadratic function. Now Proposition 1 of the preceding section applies. $\square$

4. Proof of Theorem 2.

LEMMA 1. *Let T be a bounded linear transformation of a Banach space X into a normed linear space Z. Then, for T to be surjective (onto Z), it suffices that there exist a mapping S, not necessarily linear or continuous, of*

Z into X and positive numbers α and β, $\alpha < 1$, such that, for all $g \in Z$,

$$\|g - TSg\| \leq \alpha\|g\|, \tag{1}$$

and

$$\|Sg\| \leq \beta\|g\|. \tag{2}$$

Proof of Lemma 1. Fix a $g \in Z$ and define a mapping $U = U_g$ of X into X, thus.

$$Ux = x + S(g - Tx). \tag{3}$$

Since $U^{n+1}x = U^n x + S(g - TU^n x)$ and, since T is linear,

$$\begin{aligned} \|g - TU^{n+1}x\| &= \|g - TU^n x - TS(g - TU^n x)\| \\ &\leq \alpha\|g - TU^n x\| \leq \alpha^{n+1}\|g - Tx\|, \end{aligned} \tag{4}$$

where (1) and an induction have been used. Consequently, for all x,

$$\lim TU^n x = g. \tag{5}$$

Next, use (3) and (2) to obtain

$$\|Ux - x\| = \|S(g - Tx)\| \leq \beta\|g - Tx\|. \tag{6}$$

Now use (6) and (4) to get

$$\|U^{n+1}x - U^n x\| \leq \beta\|g - TU^n x\| \leq \beta\alpha^n\|g - Tx\|. \tag{7}$$

Therefore, $\lim U^n x$ exists and

$$T(\lim U^n x) = \lim TU^n x = g \tag{8}$$

where the continuity of T and (5) have been used. □

Define a mapping S of $C(\mathcal{P})$ into $C(K^\infty)$ by letting $S(g)$ be $S_n g = g(D_n)$ when $n = n(g)$ is minimal with the property

$$\|g - TS_n g\| \leq \tfrac{1}{2}\|g\|. \tag{9}$$

That S is well defined is implied by Theorem 1. Plainly, for each $\omega \in K^\infty$, $(S_n g)(\omega) = g(D_n(\omega))$, so the range of $S_n g$ is a subset of the range of g which implies that

$$\|S_n g\| \leq \|g\|. \tag{10}$$

Since (9) and (10) plainly hold with S_n replaced by S, the condition of Lemma 1 obtains with $X = C(K^\infty)$, $Z = C(\mathcal{P})$, $\alpha = \frac{1}{2}$ and $\beta = 1$. □

Further acknowledgements. In addition to Benjamin Weiss, I am indebted to David Aldous, Paul Chernoff, Michael Fochler, Avram Neyman and Gideon Schwartz for various helpful conversations and suggestions.

REFERENCES

[1] Heinze Bauer, *Approximation and abstract boundaries*, Amer. Math. Monthly, **85**, No. 8, (1978), 632–647.

[2] E. Hewitt and L. J. Savage, *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc., **80** (1955), 470–501.

[3] P. P. Korovkin, *Linear Operators and Approximation Theory*, Hindustan Publishing Corporation, Delhi, India, (1960).

Received November 19, 1981. Research supported by NSF Grant No. MCS-80-02535.

UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720